



Tersedia online di www.journal.unipdu.ac.id
Unipdu
Terakreditasi Sinta S5

Halaman jurnal di www.journal.unipdu.ac.id/index.php/teknologi



Klasifikasi citra makanan/nonmakanan menggunakan metode *Transfer Learning* dengan model *Residual Network*

Classification of food/non-food images using Transfer Learning method with Residual Network model

Gustavo Thiodorus ^{a,*}, Anugrah Prasetya ^b, Luthfi Afrizal Ardhani ^c, Novanto Yudistira ^d

^{a,b,c,d} Teknik Informatika, Universitas Brawijaya, Malang, Indonesia

email: ^{a,*} gwa1999@student.ub.ac.id, ^b anugrahp99@student.ub.ac.id, ^d yudistira@ub.ac.id

* Koresponden

INFO ARTIKEL

Sejarah artikel:

Menerima 22 April 2021

Revisi 11 Mei 2021

Diterima 11 Mei 2021

Online 25 Agustus 2021

Kata kunci:

AlexNet

citra makanan

fine tuning

freeze layer

ResNet18

transfer learning

Keywords:

AlexNet

fine tuning

food image

freeze layer

ResNet18

transfer learning

Style APA dalam menyitasi artikel ini:

Thiodorus, G., Prasetya, A., Ardhani, L. A., & Yudistira, N. (2021). Klasifikasi citra makanan/nonmakanan menggunakan metode *Transfer Learning* dengan model *Residual Network*. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 11(2), 74-83.

ABSTRAK

Aktivitas masyarakat yang sering mengunggah foto makanan sebelum disantap lumrah dijumpai dalam media sosial populer, seperti Facebook, Instagram, Twitter, maupun Pinterest. Data citra makanan yang beredar di media sosial tersebut dapat dimanfaatkan untuk keperluan bisnis, seperti analisa pola perilaku pelanggan. Namun, tidak semua citra yang diunggah merupakan citra makanan. Sehingga, sebelum dilakukan analisis maka perlu dilakukan proses klasifikasi terlebih dahulu antara citra makanan dan citra nonmakanan. Oleh karena itu, peneliti ini mengusulkan klasifikasi citra makanan secara otomatis menggunakan metode *Transfer Learning* menggunakan model *Residual Network* versi 18 (ResNet-18). Model *Residual Network* digunakan karena mempunyai mekanisme *residual connection* untuk mengatasi masalah *vanishing gradient*. Selain itu, *transfer learning* sendiri dipilih karena metode ini mampu memanfaatkan fitur dan bobot yang telah didapatkan pada proses pelatihan sebelumnya, sehingga dapat mengurangi waktu komputasi dan meningkatkan akurasi. Pengujian dan membandingkan kemampuan model ResNet18 dan AlexNet dilakukan dalam penelitian ini. Metode *fine tuning* dan *freeze layer* untuk meningkatkan ragam kualitas model yang dihasilkan juga dilakukan pada penelitian ini. Pada penelitian ini, 3.000 citra untuk data latih, 1.000 citra untuk data uji, dan evaluasi yang digunakan adalah akurasi. Hasil yang didapatkan, yaitu model ResNet18 dengan metode *fine tuning* menghasilkan nilai akurasi sebesar 0,981 dan dengan metode *freeze layer* menghasilkan nilai terbaik akurasi sebesar 0,988. Model AlexNet dengan metode *fine tuning* menghasilkan nilai akurasi sebesar 0,970 dan metode *freeze layer* menghasilkan nilai akurasi terbaik sebesar 0,978. Hasil penelitian ini dapat disimpulkan bahwa mekanisme dengan akurasi yang paling baik terdapat pada arsitektur ResNet18 mekanisme *freeze layer* 1-3 dengan menghasilkan akurasi sebesar 0,988.

ABSTRACT

People's activities that often upload photos of food before eating are common in popular social media, such as Facebook, Instagram, Twitter, and Pinterest. The food image data circulating on social media can be used for business purposes, such as analyzing customer behavior patterns. However, not all of the uploaded images are food images, so that before the analysis is carried out, it is necessary to do a classification task between food images and non-food images. Therefore, the researcher proposes automatic food image classification using transfer learning method using the Residual Network model version of 18 (ResNet-18). Residual Network model is used because it has a residual connection mechanism to solve the vanishing gradient problem. In addition, transfer learning was chosen because this method leverages the features and weights that have been generated in

the previous training process on large and more general data (Imagenet) and thus reduce computation time and increase accuracy. The test was carried out by comparing the capabilities of the ResNet18 model with AlexNet. In addition, the fine tuning and freeze layer methods used to improve the quality of the model were also carried out in this study. In the experiment, the data set was divided into 3,000 images for training data and 1,000 images for test data, while the evaluation used was correctness accuracy. The results obtained in the ResNet18 model, namely the fine tuning training method, produced an accuracy value of 0.981 while the freeze layer resulted in the best accuracy value of 0.988. The AlexNet model that uses the fine tuning training method produces an accuracy value of 0.970 while the freeze layer produces the best accuracy value of 0.978. It can be concluded that the mechanism with the best accuracy is found in the ResNet18 architecture using the freeze layer 1-3 with an accuracy of 0.988.

Teknologi: Jurnal Ilmiah Sistem Informasi dengan lisensi CC BY NC SA.

1. Pendahuluan

Makanan merupakan suatu hal yang tidak dapat kita pisahkan dalam kehidupan sehari-hari. Setiap hari kita pasti mengonsumsi suatu makanan, entah itu sebagai camilan atau kebutuhan hidup. Apalagi di zaman yang serba instan dan canggih, banyak orang yang memiliki ide kreatif untuk mencoba dan menciptakan berbagai macam makanan yang unik, cantik, dan lezat sehingga menjadikan aneka jenis makanan semakin bervariasi. Hal tersebut dapat kita rasakan dengan munculnya banyak kedai, restoran, rumah makan, hingga pedagang kaki lima di bahu jalan taman kota.

Pada saat ini, hampir setiap orang memiliki *smartphone* yang dilengkapi dengan fitur kamera di dalamnya. Sehingga, hal ini dapat memudahkan siapa saja untuk memotret dan mengunggahnya ke dalam sosial media, salah satunya foto makanan. Foto makanan merupakan salah satu jenis foto yang sangat populer di media social (Atanasova, 2016). Hal tersebut pada saat ini banyak dijumpai unggahan foto-foto makanan di berbagai sosial media populer seperti Instagram, Twitter, Pinterest, dan Facebook. Masyarakat melihat bahwa makanan itu merupakan gaya hidup dan budaya, sehingga foto makanan tersebut menjadi suatu hal yang tren pada saat ini (Bestari, 2014).

Data citra yang terdapat pada media sosial dapat dimanfaatkan untuk kepentingan analisis bisnis, contohnya analisis pola perilaku pelanggan melalui jenis foto yang diunggah. Namun, sebelum dilakukan proses analisis bisnis perlu dilakukan klasifikasi citra untuk membedakan kategori citra yang diunggah. Salah satu jenis kategori yang dapat diklasifikasikan adalah antara makanan atau nonmakanan. Pengetahuan terkait kategori dari citra yang diunggah bermanfaat bagi para pemangku kepentingan untuk dianalisis secara lebih dalam.

Salah satu metode untuk mengklasifikasikan antara citra makanan dan nonmakanan dapat menggunakan model *deep learning* dengan metode transfer learning. Sebelumnya telah dilakukan penelitian mengenai klasifikasi makanan menggunakan model *deep learning* dengan metode *transfer learning* (Islam, Siddique, Rahman, & Jabid, 2018). Penelitian tersebut membahas mengenai komparasi akurasi setiap model *transfer learning* yang digunakan untuk klasifikasi. Pada model AlexNet mendapatkan akurasi sebesar 0,9519; GoogLeNet 0,9920; dan AlexNet-ANN 0,9940. Penelitian selanjutnya adalah mengenai klasifikasi batik menggunakan *Deep Convolutional Neural Network transfer learning* (Gultom, Arymurthy, & Masikome, 2018). Jumlah data sebanyak 2,092 potongan foto Batik (5 kelas), model yang menggunakan ConvNet *Visual Geometry Group 16* (VGG16) sebagai ekstraktor fitur mencapai rata-rata akurasi 0,897, sedangkan model berbasis *e Scale-Invariant Feature Transform* (SIFT) dan *Speeded up Robust Features* (SURF) mencapai rata-rata 0,8810 dan 0,8818. Penelitian mengenai klasifikasi makanan dan nonmakanan pun pernah dilakukan oleh Islam, Wijewickrema, Pervez, dan O'Leary (2018). Pada penelitian tersebut menggunakan *dataset* Food-11. Islam, Wijewickrema, Pervez, dan O'Leary (2018) menggunakan metode *Convolutional Neural Network* (CNN) dan *pre-trained inception V3 CNN* untuk melakukan klasifikasi citra makanan. Pada model CNN yang diajukan memperoleh nilai akurasi sebesar 0,7470 dan model *pre-trained Inception V3* 0,9286. Singla, Yuan, dan Ebrahimi (2016) juga pernah melakukan penelitian terkait klasifikasi citra makanan/nonmakanan dan pengenalan kategori makanan menggunakan model *pre-trained GoogleNet*. Penelitian tersebut menggunakan *dataset* yang dikumpulkan sendiri dari sosial media, *dataset* yang sudah ada, kamera, dan *smartphone*. Pada penelitian tersebut, model *pre-trained GoogleNet* menghasilkan nilai akurasi sebesar 0,9920 pada klasifikasi citra makanan/nonmakanan dan 0,8360 pada pengenalan kategori makanan.

Berdasarkan latar belakang yang sudah dipaparkan, maka dipilih metode *transfer learning* untuk mengklasifikasi makanan dan nonmakanan. Metode *transfer learning* merupakan metode dalam *deep neural net* yang menggunakan bobot dan fitur yang sudah dilatih menggunakan *dataset* yang besar, sehingga kaya akan fitur dan dapat mempercepat waktu komputasi. Kelebihan metode CNN dapat digunakan untuk mempelajari secara mandiri fitur-fitur pada citra yang sangat kompleks dan tidak memerlukan ekstraksi fitur apapun melainkan citra mentah (Yudistira, Widodo, & Rahayudi, 2020; Rohim, Sari, & Tibyani, 2019). Selain itu, dalam penelitian ini untuk mekanisme pengujian akan menggunakan mekanisme *fine tuning* dan *freeze layer* untuk meningkatkan ragam kualitas model yang dihasilkan dengan melihat nilai akurasi model menggunakan semua *layer* ekstraksi fitur (*fine tuning*) dan selanjutnya secara *incremental* melakukan *freeze layer* terhadap *layer* ekstraksi fitur.

2. State of the Art

Islam, Wijewickrema, Pervez, dan O'Leary (2018) melakukan klasifikasi citra makanan menggunakan *pre-trained Deep Convolutional Neural Networks* (DCNN) menggunakan *dataset* Food-5K metode DCNN dengan dua cara. Pertama, peneliti menggunakan *transfer learning* dan melatih kembali CNN pada citra makanan. Kedua, peneliti mengekstrak fitur dari CNN yang telah dilatih sebelumnya untuk melatih pengklasifikasi konvensional. Pada kedua skenario tersebut memberikan hasil tingkat akurasi yang hampir sama dan waktu pelatihan yang cukup singkat. Peneliti juga membandingkan antara metode yang diajukan dengan beberapa metode-metode klasifikasi citra yang sudah ada. Pada model AlexNet mendapatkan akurasi sebesar 0,9519, GoogLeNet 0,9920, dan AlexNet-ANN (metode yang diajukan) 0,9940.

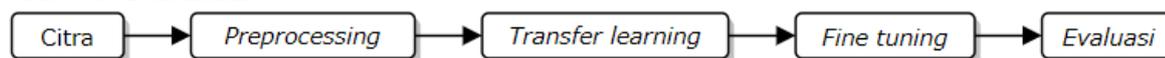
Gultom, Arymurthy, dan Masikome (2018) melakukan penelitian terkait klasifikasi batik menggunakan *Deep Convolutional Neural Network* dengan metode *transfer learning*. Penelitian ini menggunakan 2.092 citra dengan lima kelas. Lima kelas tersebut adalah Parang, Kawung, Ceplok, Nitik, dan Lereng. Berdasarkan hasil uji coba didapatkan tingkat akurasi sebesar 0,897 menggunakan model ConvNet VGG16 sebagai ekstraktor fitur; 0,8810 menggunakan model berbasis SIFT; dan 0,8880 menggunakan model berbasis SURF.

Penelitian mengenai klasifikasi makanan dan nonmakanan pun pernah dilakukan oleh Islam, Siddique, Rahman, dan Jabid (2018). Pada penelitian tersebut menggunakan *dataset* Food-11. Islam, Siddique, Rahman, dan Jabid (2018) menggunakan metode CNN dan *pre-trained inception V3* CNN untuk melakukan klasifikasi citra makanan. Pada model CNN yang diajukan didapatkan nilai akurasi sebesar 0,7470 dan model *pre-trained Inception V3* 0,9286.

Singla, Yuan, dan Ebrahimi (2016) dalam penelitiannya melakukan klasifikasi citra makanan/nonmakanan dan pengenalan kategori makanan menggunakan model *pre-trained* GoogleNet. Penelitian tersebut menggunakan *dataset* yang dikumpulkan sendiri dari sosial media, *dataset* yang sudah ada, kamera, dan *smartphone*. Pada penelitian tersebut, model *pre-trained* GoogleNet menghasilkan nilai akurasi sebesar 0,9920 pada klasifikasi citra makanan/nonmakanan dan 0,8360 pada pengenalan kategori makanan.

Rohim, Sari, dan Tibyani (2019) dalam penelitiannya melakukan implementasi metode *Convolutional Neural Network* (CNN) untuk mengklasifikasikan citra makanan tradisional. Pada arsitektur CNN yang digunakan membutuhkan 4 *layer* konvolusional, 4 *layer* *maxpooling*, dan 2 *layer* *fully connected*. Berdasarkan hasil pengujian, arsitektur tersebut mampu memperoleh nilai 0,73 presisi, 0,69 *recall*, dan 0,69 *F-score*.

3. Metode Penelitian



Gambar 1. Alur metode penelitian

Pada penelitian ini akan dilakukan klasifikasi citra makanan yang terdiri dari dua kelas (*binary classification*) menggunakan metode *transfer learning*. Metode atau alur penelitian yang dilakukan pada penelitian ini dibagi menjadi beberapa tahapan, yaitu *input* citra, *preprocessing*, *transfer learning*, *fine tuning*, dan evaluasi. Diagram alir dari metode penelitian dapat dilihat pada Gambar 1.

3.1. Citra

Pada penelitian ini digunakan objek citra makanan menggunakan *dataset* Food-5k dari situs Kaggle yang dapat diunduh di alamat <http://www.kaggle.com/trolukovich/food5k-image-dataset>. *Dataset*

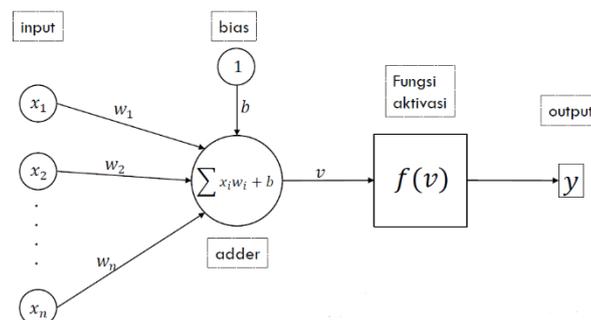
terdiri dari dua kelas yaitu makanan (*food*) dan bukan makanan (*non-food*) dengan tiga pembagian data, yaitu data evaluasi dengan 500 kelas makanan, dan 500 kelas bukan makanan. Data latih dengan 1.500 kelas makanan, dan 1.500 kelas bukan makanan. Terakhir, yaitu data validasi dengan 500 data makanan, dan 500 data bukan makanan. Total citra yang terdapat pada *dataset* Food-5k adalah 5.000 citra. Beberapa sampel dari citra pada *dataset* dapat dilihat pada Gambar 2.



Gambar 2. Contoh sampel citra dari *dataset* (Atas: Makanan, Bawah: Bukan makanan)

3.2. Jaringan Syaraf Tiruan

Jaringan Syaraf Tiruan (JST) atau *Artificial Neural Network* (ANN) merupakan model komputasi yang terinspirasi dari cara kerja otak manusia dalam memproses informasi. Konsep dari JST merupakan generalisasi model matematika dari bagaimana manusia memproses informasi pada suatu elemen sederhana yang disebut neuron, setiap neuron mempunyai penghubung tempat isyarat mengalir, setiap penghubung mempunyai bobot yang bersesuaian, dan setiap neuron akan bermuara kepada sel syaraf lainnya di mana terdapat fungsi aktivasi terhadap masukan berupa hasil penjumlahan isyarat berbobot untuk menentukan isyarat keluarannya. Selain itu pada setiap lapisan JST terdapat bias, yaitu neuron yang selalu bernilai 1. Gambar 3 merupakan struktur JST.



Gambar 3. Struktur JST

Formula v dan formula y yang disajikan melalui Gambar 3 pada JST disajikan pada Persamaan 1 dan Persamaan 2. v merupakan formula perhitungan untuk menjumlahkan seluruh *input* dan bias, sesuai dengan bobot masing-masing yang disajikan pada Persamaan 1. Lalu, untuk selanjutnya hasil perhitungan tersebut akan diaktivasi menggunakan fungsi aktivasi $f(v)$ seperti Persamaan 2 untuk mendapatkan *output* y . x adalah *input* citra, w adalah bobot, b adalah bias.

$$v = \sum_{i=1}^n x_i w_i + b \quad (1)$$

$$y = f(v) \quad (2)$$

3.3. Convolutional Neural Network

Convolutional Neural Network (CNN atau ConvNet) adalah salah satu bentuk arsitektur *deep neural network* yang dibangun dengan lapisan dasar dengan operasi konvolusi. Operasi konvolusi dilakukan sama dengan ekstraksi fitur yang ada dalam pembelajaran mesin konvensional. Perbedaannya adalah pada *deep neural network* bobot atau nilai matriks *filter* akan dicari menggunakan metode optimasi secara pembelajaran mesin tanpa campur tangan manusia. ConvNet sendiri mampu menangkap dependensi Spasial dan Temporal dalam sebuah citra melalui aplikasi *filter* yang relevan. Persamaan 3 merupakan formula perhitungan hasil konvolusi $S(i,j)$, di mana S adalah hasil konvolusi, K adalah kernel *filter, dan i adalah masukan.*

$$S(i,j) = (K \times I)(i,j) = \sum_m \sum_n i(i-m, j-n) K(m,n) \quad (3)$$

3.3.1. Fungsi aktivasi

Fungsi aktivasi adalah fungsi yang digunakan untuk mengubah keluaran dari lapisan operasi seperti konvolusi ataupun *linear* menjadi *non-linear*. Beberapa fungsi aktivasi adalah:

- Fungsi aktivasi ReLU

Fungsi aktivasi *Rectified Linear Unit* (ReLU) adalah fungsi yang berfungsi untuk menonaktifkan neuron dengan nilai keluaran negatif. Selain itu ReLU juga digunakan untuk mengatasi *vanishing gradient* karena gradien tidak dilakukan operasi ketika nilai keluaran lebih dari nol. Persamaan 4 merupakan formula perhitungan fungsi aktivasi ReLU, di mana x adalah input fungsi aktivasi.

$$R(x) = \max(0, x) \quad (4)$$

- Fungsi aktivasi SoftMax

Fungsi aktivasi SoftMax digunakan untuk mengubah *output* dari lapisan *fully connected* yang berupa vektor angka menjadi vektor probabilitas, di mana probabilitas setiap nilai sebanding dengan skala relatif setiap nilai dalam vektor. Keluaran dari fungsi aktivasi SoftMax akan digunakan untuk perhitungan nilai *loss*. Persamaan 5 merupakan formula perhitungan fungsi aktivasi SoftMax, di mana e adalah nilai Euler, z adalah *input*, dan K adalah jumlah target.

$$\text{Softmax} = e^{z_j} / (\sum_{k=1}^K e^{z_k}) \quad (5)$$

3.3.2. Loss function

Loss function adalah fungsi yang menghitung besarnya selisih prediksi dari sebuah model terhadap data aslinya. *Loss function* dipilih atau didefinisikan oleh peneliti sesuai tujuan dari pelatihan model. Dalam kasus klasifikasi, *loss function* yang sering digunakan adalah *log (cross-entropy) loss* (Janocha & Czarnecki, 2016). Fungsi *cross-entropy loss* dapat dituliskan pada Persamaan 6, di mana L adalah *loss function*, y adalah kelas target, dan o adalah keluaran model.

$$L = \sum_j y^{(j)} \log \sigma(o)^{(j)} \quad (6)$$

3.3.3. Optimasi Gradient Descent

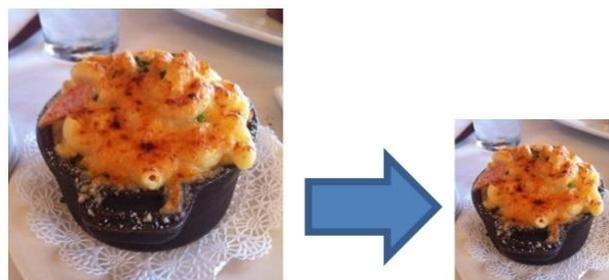
Optimisasi *gradient descent* adalah metode pembaruan bobot dengan mengevaluasi gradien performa model terhadap *fungsi loss*. Setelah gradien dievaluasi, bobot pada model diperbarui menggunakan algoritma tertentu. Algoritma yang sering digunakan dalam pelatihan adalah *Stochastic Gradient Descent* (SGD).

SGD adalah metode pembaruan bobot yang dilakukan di setiap baris data. SGD melakukan pembaruan bobot secara berkala sehingga bobot akan terus diperbarui sehingga dapat gradien dapat berkurang secara cepat. Namun, SGD memerlukan konfigurasi yang cocok untuk mengatasi sulitnya konvergen saat model mendekati konvergen (Ruder, 2016). Persamaan 7 merupakan formula pembaharuan bobot pada SGD, di mana θ adalah bobot dan η adalah *learning rate*.

$$\theta = \theta - \eta \cdot \Delta \theta J(\theta; x^{(i)}; x^{(j)}) \quad (7)$$

3.4. Preprocessing

Citra masukan yang akan dilatih dan validasi perlu dilakukan *preprocessing* terlebih dahulu. *Preprocessing* yang akan dilakukan adalah *scaling* atau merubah resolusi skala menjadi 224×224 *pixel*. *Scaling* merupakan sebuah standar *input* untuk proses CNN menggunakan metode *transfer learning* dengan arsitektur VGG16 (Swastika, 2020). Contoh *scaling* pada penelitian ini dapat dilihat pada Gambar 4.



Gambar 4. Scaling citra asli menjadi 224×224 *pixel*

Selain itu, pada data latih diterapkan metode augmentasi data menggunakan metode *flip horizontal* untuk menambah keragaman pada *dataset*. Proses normalisasi pada *dataset* juga diterapkan untuk menyeragamkan intensitas *pixel*.

3.5. Transfer Learning

3.5.1. ResNet-18

ResNet-18 atau *Residual Neural Network* 18 merupakan salah satu arsitektur yang ada pada *deep learning* (He, Zhang, Ren, & Sun, 2016). ResNet mempunyai mekanisme *residual connection* yang merupakan bentuk koneksi pada JST yang dibuat dengan menambahkan jalan pintas di antara dua titik. Koneksi *shortcut* yang ada pada *residual network* ditujukan untuk mengatasi masalah *vanishing gradient* (He, Zhang, Ren, & Sun, 2016). *Vanishing gradient* adalah keadaan ketika jaringan terlalu dalam, maka gradien dari fungsi *loss* menyusut mendekati nilai 0 setelah beberapa iterasi sebelum mencapai konvergen. Akibatnya, dalam proses pelatihan bobot-bobot pada lapisan atas tidak akan pernah diperbaharui nilainya, oleh karena itu tidak ada proses pembelajaran yang dilakukan. Penambahan jalan pintas pada arsitektur ResNet ini berfungsi agar metode optimasi dapat memperbarui bobot pada gradien pada *layer* yang lebih awal sehingga *layer* awal dapat diperbarui bobotnya secara lebih baik. ResNet-18 sendiri merupakan salah satu varian dari arsitektur ResNet.

ResNet18 memiliki 18 *layer* sesuai namanya yang terdiri atas 17 *layer* konvolusi dan 1 *fully-connected layer*. Pada conv1 ResNet18 akan memiliki ukuran keluaran sebesar 112×112 . Pada conv2_x ResNet 18 akan memiliki ukuran keluaran sebesar 28×28 . Pada conv3_x ResNet 18 akan memiliki ukuran keluaran sebesar 14×14 . Pada conv4_x ResNet 18 akan memiliki ukuran keluaran sebesar 14×14 . Pada conv5_x ResNet 18 akan memiliki ukuran keluaran sebesar 7×7 . Lalu pada *layer fully-connected*, ResNet18 akan memiliki ukuran keluaran sebesar 1×1 . ResNet18 sendiri memiliki *Floating Point Operations* (FLOPs) sebanyak $1,8 \times 10^9$.

3.5.2. AlexNet

AlexNet merupakan model *deep neural network* yang ditemukan pada saat kompetisi the ImageNet *Large-Scale Visual Recognition Challenge* (ILSVRC) tahun 2012. Model ini berhasil menempati juara 1 pada kategori *lowest top-1 error rate*. AlexNet mampu memberikan hasil yang bagus pada *dataset* yang berjumlah besar dan resolusi yang sangat tinggi (Gao, 2017).

AlexNet terdiri atas delapan *layer*, yaitu lima *layer* konvolusi dan 3 *fully connected layer*. Masalah yang diselesaikan pada metode AlexNet sendiri adalah keterbatasan model lain dalam memproses *dataset* yang berjumlah sangat besar dan citra dengan resolusi yang sangat tinggi. AlexNet menyelesaikan masalah tersebut dengan membagi neuronnya pada dua GPU. Sehingga prosesnya dapat dikerjakan secara paralel. Selain itu AlexNet juga memiliki kelebihan karena telah menggunakan aktivasi RELU, dimana fungsi aktivasi tanh adalah fungsi aktivasi yang standar pada masa itu. sebuah CNN yang menggunakan RELU mampu mencapai kesalahan 25% pada set data CIFAR-10 enam kali lebih cepat daripada CNN yang menggunakan tanh (Krizhevsky, Sutskever, & Hinton, 2017). AlexNet juga menggunakan teknik *overlapping pooling*, di mana teknik ini mampu menurunkan kesalahan sebesar 0,5% (Krizhevsky, Sutskever, & Hinton, 2017).

3.6. Fine Tuning

Fine tuning merupakan teknik atau istilah yang digunakan untuk melatih kembali model yang digunakan pada proses *transfer learning*. Secara umum, hanya lapisan *fully connected* saja yang dilakukan proses *fine tuning* dengan mengganti *output* neuron sesuai dengan jumlah kelas pada *dataset* yang digunakan. Namun, apabila *dataset* yang digunakan jauh berbeda dengan *dataset* ImageNet (*pre-trained*) dapat dilakukan proses *fine tuning* pada lapisan-lapisan atas yang bertujuan untuk menyesuaikan model dengan *dataset* yang digunakan, sehingga fitur-fitur yang didapatkan bersifat lebih spesifik terhadap *dataset*. Pada penelitian ini akan berfokus untuk meneliti pengaruh *fine tuning* pada *layer* yang terdapat pada arsitektur ResNet-18 dan AlexNet terhadap akurasi yang didapatkan. Proses *fine tuning* dilakukan dengan iterasi selama 25 epoch menggunakan optimasi *stochastic gradient descent* (SGD) dan *learning rate* 0,001. Selain itu, ukuran *batch* yang digunakan adalah 4 yang bertujuan untuk menghemat penggunaan memori selama proses *fine tuning*. Fungsi *loss* yang digunakan adalah *cross entropy* yang memperhitungkan hasil *output* neuron dengan normalisasi *softmax* dan jarak terhadap label.

3.7. Freeze Layer

Freeze layer merupakan salah satu teknik yang digunakan pada proses *fine tuning*. *Freeze layer* merupakan teknik yang digunakan untuk mengontrol bagaimana bobot pada *layer* diperbaharui. Ketika

layer tersebut dibekukan, maka bobot pada *layer* tersebut tidak akan diperbaharui. Sebaliknya, ketika *layer* tersebut tidak dibekukan maka proses pembaharuan bobot akan terus dilakukan pada *layer* tersebut. Pada penelitian kali ini dilakukan beberapa skenario *freeze layer* pada arsitektur ResNet-18 dan AlexNet. Skenario tersebut diantaranya adalah *freeze conv 1-4*, *freeze conv 1-3*, *freeze conv 1-2*, *freeze conv 1*, dan *freeze* semua *layer* konvolusi terkecuali lapisan *fully connected*. Skenario tersebut akan menganalisis hasilnya dari masing-masing skenario dan pengaruhnya terhadap akurasi.

3.8. Evaluasi

Evaluasi performa dari metode yang diusulkan menggunakan akurasi. Akurasi sendiri merupakan hasil bagi antara jumlah data yang masuk di kelas yang tepat dibagi dengan total jumlah semua data. Untuk *dataset* sendiri kami bagi menjadi 80% untuk data latih dan 20% untuk data uji.

4. Hasil dan Pembahasan

Pada bagian ini akan dijelaskan mengenai uji coba yang akan dilakukan untuk mendapatkan akurasi yang terbaik dari arsitektur dan mekanisme terbaik. Mekanisme yang akan dijadikan percobaan adalah dilakukan *freeze* setiap *layer* di setiap arsitektur untuk mendapatkan hasil akurasi dan bisa dibandingkan. Untuk mendapatkan akurasi dengan cara menebak jumlah yang benar (*True Positive*) dibagi dengan jumlah keseluruhan data. Persamaan untuk menentukan akurasi dapat dilihat pada Persamaan 8.

$$Accuracy = \frac{True\ Positive}{Sum\ of\ Data} \quad (8)$$

4.1. Pengujian Arsitektur ResNet-18

Pada pengujian arsitektur ResNet-18 kali ini menggunakan *epoch* 25, *learning rate* 0,001, menggunakan *optimizer* SGD. Untuk mekanisme pengujian, kami menggunakan *fine tuning* pada semua *layer*, *freeze conv 1-4*, *freeze conv 1-3*, *freeze conv 1-2*, *freeze conv 1*, dan *freeze all conv*. Perbandingan hasil akurasi dapat dilihat pada Tabel 1. Grafik perbandingan antara nilai *loss* dan akurasi baik *training* maupun *validation* pada setiap skenario dapat dilihat pada Gambar 4.

Perbedaan hasil akurasi dari setiap percobaan mekanisme tidak berbeda secara signifikan. Hasil dengan akurasi tertinggi terdapat pada pengujian mekanisme *freeze conv 1-2* dengan akurasi sebesar 0,988. Hasil dengan akurasi terkecil terdapat pada pengujian mekanisme *freeze conv 1* dengan akurasi sebesar 0,979. Jika dibandingkan antara skenario pengujian *fine tuning* (tanpa *freeze*) dan *freeze conv 1*, nilai akurasi mengalami kenaikan. Jika dibandingkan antara skenario pengujian *freeze conv 1* dan *freeze conv 1-2*, nilai akurasi mengalami kenaikan dan mencapai nilai maksimum. Namun, antara skenario pengujian *freeze conv 1-2* dan *freeze conv 1-3* terjadi penurunan nilai akurasi. Kemudian, pada skenario pengujian *freeze conv 1-3*, *freeze conv 1-4*, *freeze all layers* tidak terjadi perubahan nilai akurasi. Pada pengujian ini, semua model memiliki hasil akurasi yang baik dengan selisih perbedaan yang sangat kecil. Dari fakta-fakta tersebut dapat disimpulkan bila semakin banyak *layer* yang dibekukan tidak berbanding lurus terhadap kenaikan nilai akurasi model.

Pada Gambar 4, dapat dilihat bila nilai *loss* yang dihasilkan pada masing masing skenario pengujian cenderung baik. nilai *loss* pada data *training* berkisar antara 0,1-0,3. pada data *validation*, nilai *loss* berkisar antara 0,01-0,2. Pada data *validation*, nilai *loss* terkecil dicapai pada skenario pengujian *freeze conv 1-2* dengan nilai kurang dari 0,05. Nilai *loss* terbesar dicapai pada skenario pengujian *fine tuning* dengan nilai lebih besar dari 0,15. Semakin kecil nilai *loss* menunjukkan bila model yang dihasilkan memiliki performa yang baik dan mampu melakukan klasifikasi dengan baik.

4.2. Pengujian Arsitektur AlexNet

Pada pengujian arsitektur Alexnet kali ini menggunakan *epoch* 25, *learning rate* 0,001, menggunakan *optimizer* SGD. Untuk mekanisme pengujian, kami menggunakan *fine tuning*, *freeze all conv*, *freeze conv 1-4*, *freeze conv 1-3*, *freeze conv 1-2*, *freeze conv 1*. Grafik perbandingan antara nilai *loss* dan akurasi baik *training* maupun *validation* pada setiap skenario dapat dilihat pada Gambar 5. Perbandingan hasil akurasi dapat dilihat pada Tabel 1.

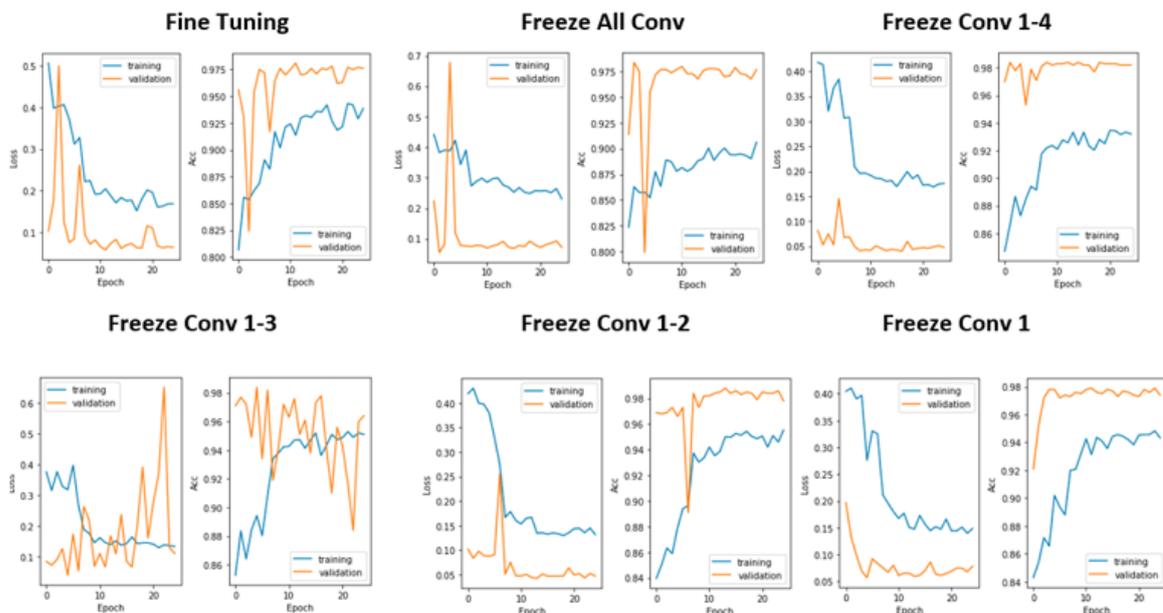
Perbedaan hasil akurasi dari setiap percobaan mekanisme tidak berbeda secara signifikan. Hasil dengan akurasi tertinggi terdapat pada pengujian mekanisme *freeze conv 1-3* dengan akurasi sebesar 0,978. Hasil dengan akurasi terkecil terdapat pada pengujian mekanisme *fine tuning* (tanpa *freeze*) dengan akurasi sebesar 0,97. Jika dibandingkan antara skenario pengujian *fine tuning* (tanpa *freeze*) dan *freeze conv 1*, nilai akurasi mengalami kenaikan sebanyak 0,005. Jika dibandingkan antara skenario

pengujian *freeze conv 1* dan *freeze conv 1-2*, nilai akurasi mengalami penurunan sebesar 0,003. Jika dibandingkan antara skenario pengujian *freeze conv 1-2* dan *freeze conv 1-3*, nilai akurasi mengalami kenaikan dan mencapai nilai maksimum. Kemudian, pada skenario pengujian *freeze conv 1-3*, *freeze conv 1-4*, *freeze all layers* tidak terjadi penurunan nilai akurasi. Pada pengujian ini, semua model memiliki hasil akurasi yang baik dengan selisih perbedaan yang sangat kecil. Dari fakta-fakta tersebut dapat disimpulkan bila semakin banyak *layer* yang dibekukan tidak berbanding lurus terhadap kenaikan nilai akurasi model.

Pada Gambar 5, dapat dilihat bila nilai *loss* yang dihasilkan pada masing masing skenario pengujian cenderung baik. nilai *loss* pada data *training* berkisar antara 0,025–0,1. pada data *validation*, nilai *loss* berkisar antara 0,025-0,15. Pada data *validation*, nilai *loss* terkecil dicapai pada skenario pengujian *freeze conv 1-2* dengan nilai kurang dari 0,05. nilai *loss* terbesar dicapai pada skenario pengujian *fine tuning* dengan nilai lebih besar dari 0,1. Semakin kecil nilai *loss* menunjukkan bila model yang dihasilkan memiliki performa yang baik dan mampu melakukan klasifikasi dengan baik.

Tabel 1. Perbandingan akurasi ResNet-18 dan AlexNet

No.	Mekanisme	Akurasi AlexNet	Akurasi ResNet
1	<i>Fine Tuning</i> (Tanpa <i>Freeze</i>)	0,970000	0,981000
2	<i>Freeze All Layers</i> (ConvNet as Fixed Feature Extractor)	0,974000	0,984000
3	<i>Freeze Conv 1-4</i>	0,977000	0,984000
4	<i>Freeze Conv1-3</i>	0,978000	0,984000
5	<i>Freeze Conv 1-2</i>	0,972000	0,988000
6	<i>Freeze Conv 1</i>	0,975000	0,979000

Gambar 4. Grafik *loss* dan akurasi setiap mekanisme pengujian ResNet-18

5. Kesimpulan

Berdasarkan penelitian yang telah kami lakukan dapat disimpulkan bahwa, proses klasifikasi citra makanan/nonmakanan dengan *dataset* Food-5K menggunakan metode *transfer learning* dengan model arsitektur ResNet18 dan dilakukan *freeze layer* pada setiap *layer* menghasilkan akurasi yang bervariasi. Pada mekanisme *fine tuning* menghasilkan akurasi sebesar 96,6%, *freeze conv 1-4* 98,2%, *freeze conv 1-3* 98,3%, *freeze conv 1-2* 97,9%, *freeze conv 1* 98,2%, dan *freeze all conv* menghasilkan 97,9%.

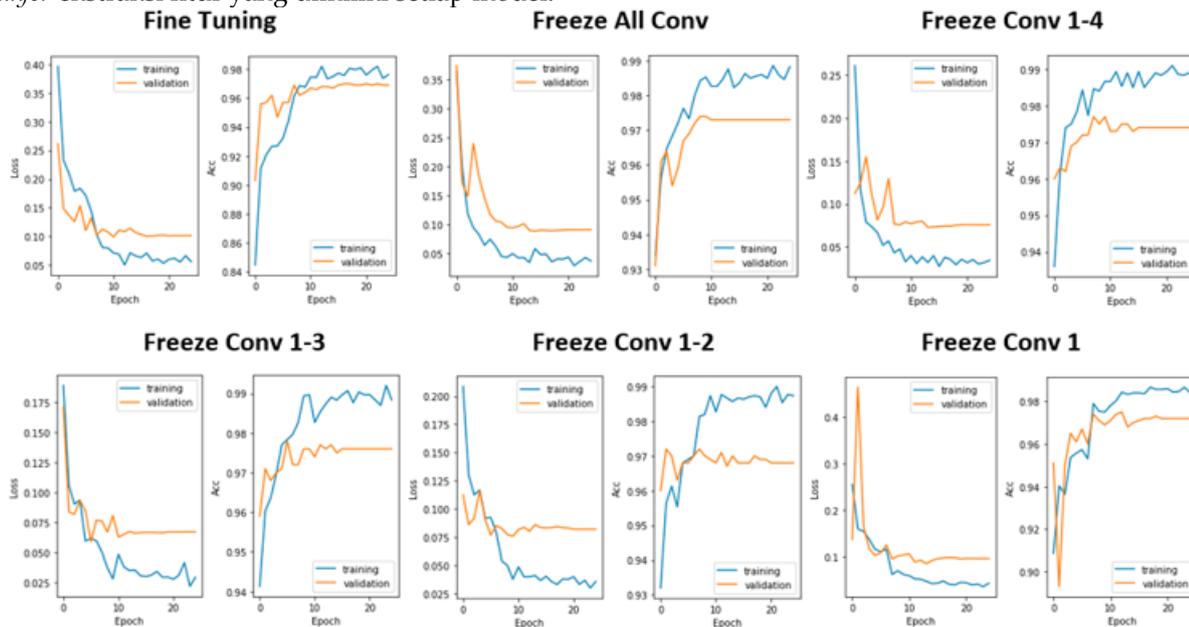
Pada arsitektur AlexNet dilakukan *freeze conv* pada setiap *layer* dan menghasilkan akurasi yang bervariasi. pada mekanisme *fine tuning* menghasilkan akurasi sebesar 97,0%, *freeze conv 1-5* 97,4%, *freeze conv 1-4* 97,7%, *freeze conv 1-3* 97,8%, *freeze conv 1-2* 97,2%, dan *freeze conv 1* 97,5%. Dari berbagai skenario pengujian yang dilakukan dengan arsitektur ResNet18 dan AlexNet dapat disimpulkan, semakin banyak *layer* yang dibekukan tidak berbanding lurus terhadap kenaikan nilai akurasi.

Perbandingan metode pengujian menggunakan arsitektur ResNet18 dan AlexNet pada metode *fine tuning* dengan mekanisme *freeze* pada setiap *layer* menghasilkan akurasi yang tinggi. pada metode

arsitektur ResNet18 memiliki akurasi tertinggi pada mekanisme *freeze layer* 1-4 dengan perolehan akurasi sebesar 98,8%. sedangkan pada arsitektur AlexNet memiliki akurasi tertinggi pada mekanisme *freeze conv* 1-3 dengan perolehan akurasi sebesar 97,8%.

Dari semua mekanisme tersebut dapat disimpulkan bahwa, mekanisme dengan akurasi yang paling baik terdapat pada arsitektur ResNet18 mekanisme *freeze layer* 1-2 dengan menghasilkan akurasi sebesar 98,8%. ResNet18 memiliki mekanisme *skip connection* yang membuat dia lebih unggul daripada metode AlexNet. ResNet18 juga memiliki jumlah *layer* yang lebih sedikit jika dibandingkan dengan AlexNet. ResNet18 mampu menyelesaikan permasalahan *vanishing gradient* dengan mekanisme *skip connection* yang dimilikinya.

Penelitian selanjutnya, dapat menggunakan model-model yang lain dengan menggunakan metode *transfer learning* dengan mekanisme pengujian *fine tuning* dan *freeze layer* sesuai dengan jumlah *layer* ekstraksi fitur yang dimiliki setiap model.



Gambar 5. Grafik *loss* dan akurasi setiap pengujian AlexNet

6. Referensi

- Atanasova, A. (2016, Nov. 9). *The Psychology of Foodstagramming*. Retrieved from Social Media Today: <https://www.socialmediatoday.com/social-networks/psychology-foodstagramming>
- Bestari, K. (2014). *Fenomena Mengunggah Foto Makanan pada Pengguna Media Sosial*. Depok: Universitas Indonesia.
- Gao, H. (2017, Aug. 8). *A Walk-through of AlexNet*. Retrieved from Hao Gao: <https://medium.com/@smallfishbigsea/a-walk-through-of-alexnet-6cbd137a5637>
- Gultom, Y., Arymurthy, A. M., & Masikome, R. J. (2018). Batik Classification using Deep Convolutional Network Transfer Learning. *Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)*, 11(2), 59-66.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). Las Vegas, NV, USA: IEEE.
- Islam, K. T., Wijewickrema, S., Pervez, M., & O'Leary, S. (2018). An Exploration of Deep Transfer Learning for Food Image Classification. *Digital Image Computing: Techniques and Applications (DICTA)*. Canberra, ACT, Australia: IEEE.
- Islam, M. T., Siddique, B. M., Rahman, S., & Jabid, T. (2018). Food Image Classification with Convolutional Neural Network. *International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (pp. 257-262). Bangkok, Thailand: IEEE.
- Janocha, K., & Czarnecki, W. M. (2016). On Loss Functions for Deep Neural Networks in Classification. Retrieved from <http://arxiv.org/abs/1702.05659>

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6), 84–90.
- Rohim, A., Sari, Y. A., & Tibyani, T. (2019). Convolution Neural Network (CNN) Untuk Pengklasifikasian Citra Makanan Tradisional. *JPTIIK (Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer)*, 3(7), 7037-7042.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. Retrieved from <https://arxiv.org/abs/1609.04747>
- Singla, A., Yuan, L., & Ebrahimi, T. (2016). Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model. *MADiMa '16: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 3-11). ACM.
- Swastika, W. (2020). Studi Awal Deteksi COVID-19 Menggunakan Citra CT Berbasis Deep Learning. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), 629-634.
- Yudistira, N., Widodo, A. W., & Rahayudi, B. (2020). Deteksi COVID-19 pada Citra Sinar-X Dada menggunakan Deep Learning yang Efisien. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(6), 1289-1296.